illumina®

High-resolution microbiome profiling with metagenomics next-generation sequencing and CosmosID

Accurate, fast, and simple microbiome analysis with the CosmosID cloud-based bioinformatics platform

Highlights

- **Fast, easy bioinformatics tool**
  Simple web-based analysis software that does not require bioinformatics expertise
- **Comprehensive microbiome analysis**
  Precise strain-level classification with curated genome databases spanning all microbial kingdoms
- **Advanced visualization tools**
  Interactive data visualization with publication-quality output

Introduction

The microorganisms that live inside and on humans, collectively known as the microbiota, exist in a symbiotic relationship, providing traits that humans did not need to evolve on their own.[1] To fully understand the range of human genetic and physiological diversity, the genomes of the microbiota, ie, the microbiome, must be considered. Research initiatives such as the Human Microbiome Project (HMP) have used next-generation sequencing (NGS) methods such as 16S rRNA sequencing and metagenomics NGS (mNGS) to advance our understanding of how the microbiome impacts human health and disease.[2–3]

Unlike capillary sequencing or PCR-based approaches, mNGS enables analysis of the entire microbial community within a sample. With microbial genomics using mNGS, public health scientists can identify microbes, examine biological functions such as antimicrobial resistance, track genetic changes, rapidly respond to outbreaks, and more. The ability to analyze sequencing data to identify and characterize pathogens is crucial for success.

This application note highlights the use of CosmosID, a microbial genomics analysis platform (CosmosID Inc., Rockville, MD), as part of a comprehensive mNGS solution that includes Illumina library prep and sequencing (Figure 1). CosmosID enables identification and classification of pathogens and commensal microorganisms in microbiome samples.

Fast, easy bioinformatics with CosmosID

Sensitive and precise identification and characterization of pathogens and commensals is achieved by CosmosID with its secure and user-friendly bioinformatics platform for analysis of NGS data. CosmosID offers a simple, quick, and easy-to-use web-based application (available also as a command line interface) to deliver exceptional performance. It scales effortlessly to handle thousands of samples for mNGS and amplicon (16S and ITS) sequencing analysis.

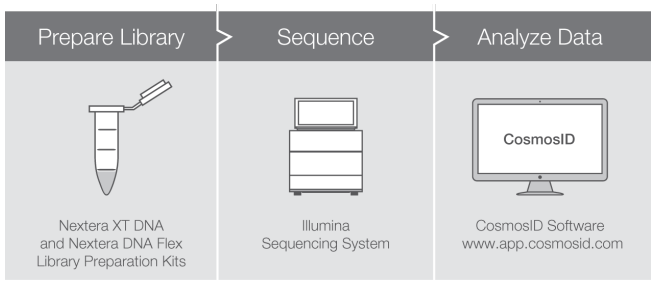

Figure 1: Metagenomics sequencing workflow with CosmosID data analysis—CosmosID, a microbial genomics platform, is part of a streamlined mNGS workflow that includes Nextera library prep and Illumina sequencing to enable rapid characterization of microorganimsms.

Experimental methods and results

To demonstrate the exceptional performance of CosmosID in microbiome analysis, mNGS and 16S rRNA metagenomic sequencing were performed on 67 fecal samples collected during a systemic surveillance at the National Institute of Cholera and Enteric Diseases (NICED) in Kolkata, India. Illumina sequencing was followed by analysis with CosmosID.

Sample collection

The 67 fecal samples collected from Kolkata, India, as part of this study were divided into three groups (Table 1). Forty-nine samples were collected from individuals affected with diarrheal disease, (17 samples with known etiology and 32 samples with unknown etiology). The cause of disease of samples with unknown etiology could not be determined using classical microbiology methods. Eighteen samples were collected from unaffected individuals for comparison. Additionally, 20 fecal metagenomic data sets obtained from the Human Microbiome Project (HMP) were included in the data analysis to serve as representation of unaffected gut microbiomes of the Western community.

Table 1: Fecal samples included in metagenomics analyses

Sample group	No. of samples
Known etiology	17
Unknown etiology	32
Unaffected control	18

Library preparation and sequencing

Libraries were prepared using the Nextera™ XT DNA Library Preparation Kit and sequenced using 1 × 100 bp reads.

Data analysis

Data analysis was performed with the CosmosID bioinformatics software package.[4–8] A key feature of CosmosID is the GenBook curated genome databases that include bacterial, fungal, and viral pathogens; commensals; environmental microorganisms; and protists. CosmosID combines GenBook databases with high-performance data-mining algorithms for multikingdom taxonomic classification of metagenomic data sets. It is important to note that GenBook databases are organized phylogenetically and contain hundreds of millions of marker sequences. The markers represent both coding and noncoding sequences uniquely identified by taxon and/or distinct nodes of phylogenetic trees.

The unassembled sequencing reads were directly analyzed to achieve multikingdom taxonomic identification to the species, subspecies, and/or strain level. Analysis also included quantification to obtain relative abundance of each taxon. Shotgun metagenome data sets were also interrogated with CosmosID resistome and virulome libraries to detect sequences encoding resistance (community resistome) and virulence factors (community virulome). This step also allowed identification of specific microbial agents within the community carrying resistance and virulence genes.

Improved microbial profiling with mNGS

Comparison of relative abundance profiles between 16S rRNA sequencing and mNGS shows improved characterization of the microbiota present in the fecal samples studied with mNGS. mNGS identified *Streptococcus* in the gut microbiome of healthy control samples and *Shigella* in the gut microbiome of both known and unknown etiology samples (Figure 2). Given the observed accuracy and speed of data processing, mNGS was used as a preferred method for metagenomic analysis of all samples in this study.

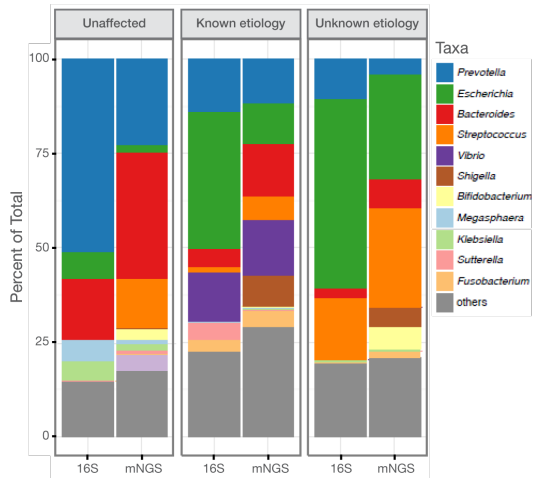

Figure 2: Comparison of 16S rRNA sequencing and mNGS—Comparison of relative abundance profiles of fecal samples shows improved performance of mNGS over 16S RNA sequencing.

Comparative analysis of gut microbiomes of affected and unaffected fecal samples

Analysis of mNGS data with CosmosID revealed significant differences in the microbial communities associated with diarrheal fecal samples (known and unknown etiology) from unaffected samples collected from individuals living in India and Western communities (HMP data). A subpopulation of the microbiome can be seen as overrepresented in samples from affected individuals, as opposed to unaffected individuals (Figure 3A). The microbiome profiles of unaffected Indian individuals were significantly different from those of unaffected Western individuals, and the former contained pathogenic microorganisms in low numbers (Figure 3). These data indicate that unaffected individuals in India may tolerate low levels of pathogens which, if present in Western individuals, would be associated with a disease state. In affected samples of known etiology, the CosmosID metagenomic analyses identified pathogens that included *Vibrio cholerae*, *Shigella spp.*, *Escherichia coli* (*E. coli*), and *Campylobacter spp.*, whereas samples from the unknown etiology group were found to contain predominantly members of the *E. coli* super family, namely pathogenic *E. coli*, *Shigella spp.*, *Salmonella enterica*, *Aeromonas caviae*, *C. jejuni*, and *Cryptosporidium spp* (Figure 3B). Multiple pathogens were often detected and affected individuals were clearly differentiated from unaffected individuals, based on microbial abundance and diversity.

Comparison of gut microbiome diversity

Population diversity analysis of HMP gut microbiome data indicates that the gut microbiomes of Western individuals are more similar in composition than the microbiomes of the Indian population, likely attributable to socioeconomic and geographical differences represented by the Indian population (Figure 4).

Characterization of antibiotic resistance profiles of gut microbiomes

Antibiotic resistome profiling enables identification of genes associated with resistance to several classes of antibiotics. Antibiotic resistance profiles were detected in individuals of all three groups (Figure 5). HMP samples did not contain genes matching at this level of coverage (> 50% gene coverage).

Functional gene profiles of gut microbiomes

Functional analysis of gut microbiomes and subsequent clustering based on Pfam and Gene Ontology (GO) provided useful insight into the role gut microorganisms may play in health and disease. In general, a predominance of genes associated with carbohydrate metabolism was noted in the gut flora of the Indian population samples. This occurence is most likely explained by the relatively carbohydrate rich diet in India, particularly the heavy consumption of rice. Compared to samples from unaffected individuals, samples from affected individuals with known or unknown etiology carried a significantly higher number of genes associated with hemolysis, cytolysis, cell killing, and bacteria-directed toxicity. Genes coding for

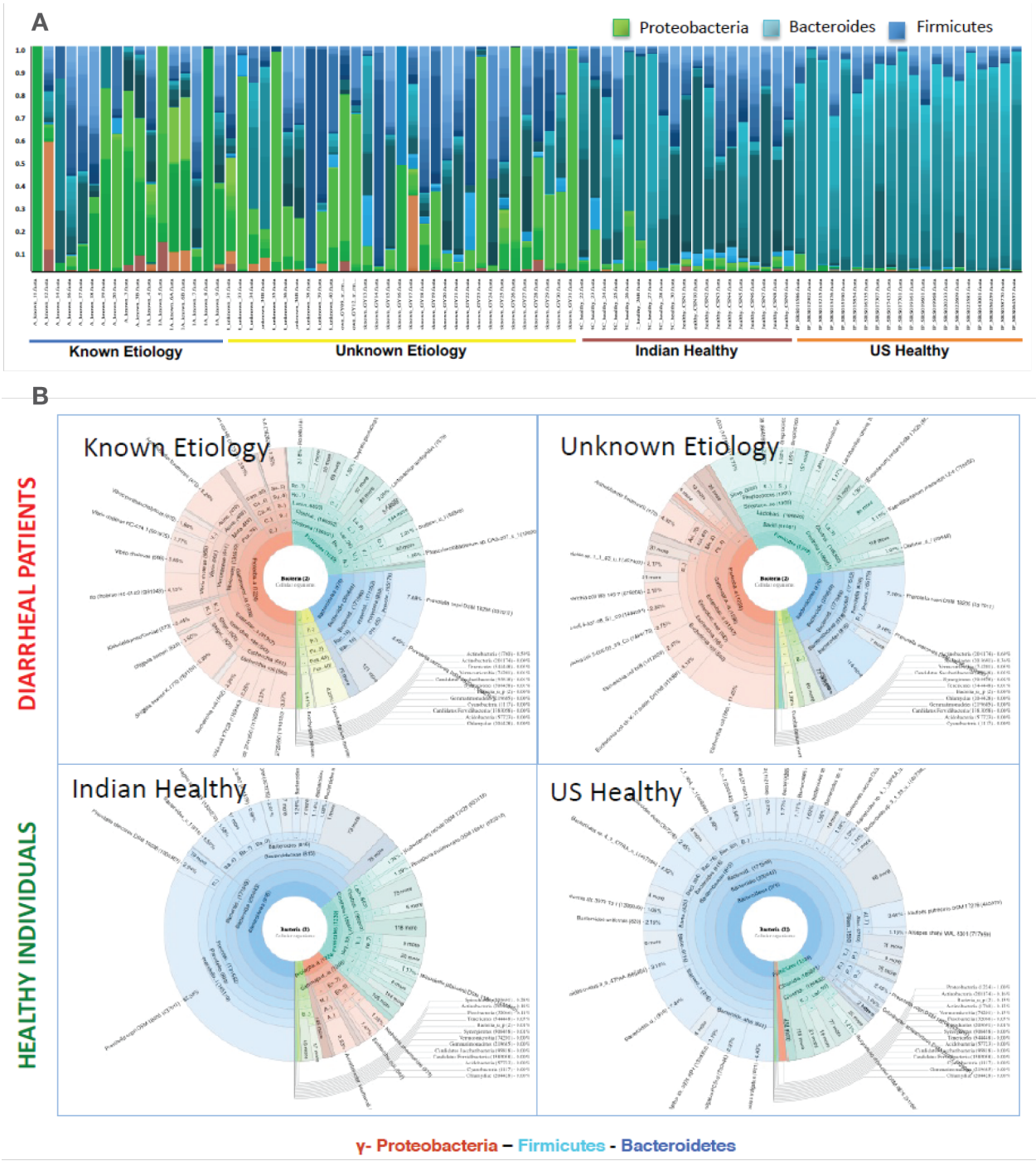

Figure 3: Comparative analysis of gut microbiomes with CosmosID—Analysis of fecal samples from affected (known and unknown etiology) individuals shows significantly different gut microbiomes from unaffected Indian individuals and Western individuals, as visualized by (A) bar chart data plot and (B) sunburst data visualization in CosmosID.

pathways of protein and DNA metabolism, synthesis of amino acids and their derivatives, and membrane transport were detected in the metagenomes of unaffected samples from Indian individuals. Gene ontology of unaffected samples also revealed a notable increase in the number of a gene sets involved in gene transfer and up-regulation of overall metabolism, biosynthesis, and transcription, compared to results obtained for the samples from Western individuals (date not shown).

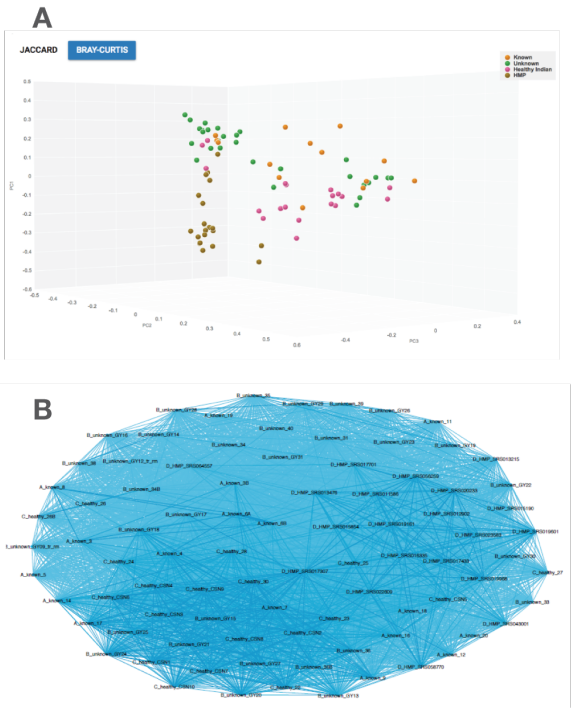

Figure 4: Comparison of gut microbiome diveristy—Population diversity analysis shows that the gut microbiomes of samples from Western individuals are more similar than samples from the Indian population, both by (A) principal component analysis and (B) beta diversity network analysis.

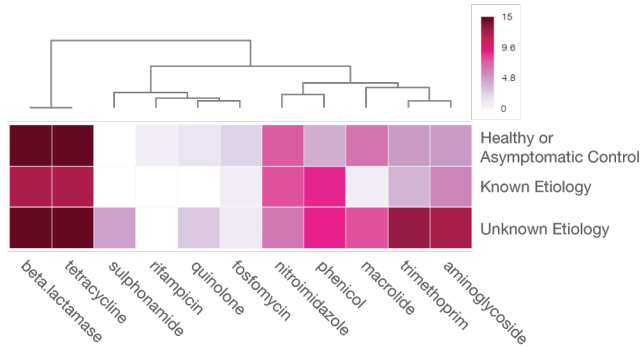

Figure 5: Characterization of antibiotic resistance—Antibiotic resistance is seen in affected samples of known and unknown etiology as well as unaffected samples.

Summary

Metagenomics sequencing is rapidly becoming the method of choice for high-resolution microbiome characterization, especially when coupled with the highly accurate and validated bioinformatics tools of the CosmosID platform. This application note demonstrates that genome profiles of gut microbial populations can be used to differentiate healthy, diseased, and asymptomatic carriers or individuals in early stages of a disease. Direct sequencing and metagenomic analysis of infectious disease samples in real time has the potential to revolutionize diagnostic, prophylactic, and therapeutic strategies.

Learn More

To learn more about CosmosID and view validation and performance metrics for the software, visit www.cosmosid.com/blog-cosmosid/benchmarking-genome-biology-2017

To analyze sequence data from example data sets, sign up for a free trial of CosmosID at app.cosmosid.com

References

1. Gill SR, Pop M, Deboy RT, et al. Metagenomic analysis of the human distal gut microbiome. *Science*. 2006;312(5778):1355–1359.
2. Turnbaugh PJ, Ley RE, Hamady M, et al. The human microbiome project: exploring the microbial part of ourselves in a changing world. *Nature*. 2007;449(7164):804–810.
3. McGregor AC, Moore DA. Infectious diseases of fever of unknown origin. *Clin Med* (Lond). 2015;15(3):285–287.
4. Connelly S, Bristol JA, Hubert S, et al. SYN-004 (ribaxamase), an oral beta-lactamase, mitigates antibiotic-mediated dysbiosis in a porcine gut microbiome model. *J Appl Microbiol*. 2017;doi:1111/jam.13432.
5. Hasan NA, Young BA, Minard-Smith AT, et al. Microbial community profiling of human saliva using shotgun metagenomic sequencing. *PLoS One*. 2014;9(5):e97699.
6. Lax S, Smith DP, Hampton-Marcell J, et al. Longitudinal analysis of microbial interaction between humans and the indoor environment. *Science*. 2014;345(6200):1048–1052.
7. Ottesen A, Ramachandran P, Reed E, et al. Enrichment dynamics of Listeria monocytogenes and the associated microbiome from naturally contaminated ice cream linked to a listeriosis outbreak. *BMC Microbiol*. 2016;16(1):275.
8. Ponnusamy D, Kozlova EV, Sha J, et al. Cross-talk among flesh-eating Aeromonas hydrophila strains in mixed infection leading to necrotizing fasciitis. *Proc Natl Acad Sci U S A*. 2016;113(3):722–727.

Illumina, Inc. • 1.800.809.4566 toll-free (US) • +1.858.202.4566 tel • techsupport@illumina.com • www.illumina.com

illumina®